EIGENVALUE CRITERION-BASED FEATURE SELECTIONIN PRINCIPAL COMPONENT ANALYSIS OF SPEECH

Peter VISZLAY[1], *Jozef JANECKO*[1], *Jozef JUHAR*[1]

[1]Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Park Komenskeho 13, 042 00 Kosice, Slovak Republic

peter.viszlay@tuke.sk, jozef.janecko@student.tuke.sk, jozef.juhar@tuke.sk

Abstract. *This article presents a specific approach for selecting a limited set of most relevant, information rich speech data from the whole amount of training data. The proposed method uses Principal Component Analysis (PCA) to optimally select a lower-dimensional data subset with similar variances. In this paper, three selection algorithms, based on eigenvalue criterion are presented. The first one operates and analyzes the data at the entire speech-recording level. The second one additionally segments each of the recordings into experimentally sized blocks, which theoretically divides a record level into several smaller information richer/poorer blocks. Finally, the third one analyzes all the speech records at the feature vector level. These three approaches represent three different criterion-based selection techniques from the coarsest to the finest data level. The main aim of the presented experiments is to show that PCA trained with the limited subset of data achieves comparable or even better results than PCA trained with the entire speech corpus. In fact, this approach can radically speed up the learning of PCA with much smaller memory and computational costs. All methods are evaluated in Slovak phoneme-based large vocabulary continuous speech recognition task.*

Keywords

Eigenvalue, feature vector, principal components, selection criterion, variance.

1. Introduction

Linear feature transformations are well-used techniques in high-dimensional data processing such as face and automatic speech recognition (ASR). The most popular transformations in automatic speech recognition are Principal Component Analysis (PCA), [1], [2], [3] and Linear Discriminant Analysis (LDA), [4]. Our speech recognition research group tends to follow the modern trends in ASR. Therefore, we are interested in research and application of linear transformations in our speech recognition system.

It is known that one integral part of PCA is the covariance matrix computing from the training set. In case of relatively small training corpus there is no problem to compute the covariance matrix. But, in case of large corpus (thousands of recordings) and high-dimensional data there may occur a problem with processing time (≈ several hours) and memory requirements (≈ 20 GB). In order to solve these problems we have built upon our previous work [5], [6] and we proposed a procedure to train PCA from a limited amount of training data. In other words, PCA can be learned from a limited subset, while the performance is maintained, or even improved. We called this procedure as *Partial-data trained* PCA. It is based on eigenvalue criterion and it is applied to LMFE (Logarithmic Mel-Filter Energies) feature vectors. The performance of the method is evaluated on Slovak speech corpus in phoneme-based continuous speech recognition task.

This paper is organized as follows. The next section gives the mathematical background of PCA. Section III describes the full-data trained PCA. Section IV presents the proposed algorithms for data selection. Section V describes the experimental setup of experiments and finally, Section VI concludes the paper.

2. Principal Component Analysis

Principal component analysis (PCA), [2] is a linear feature transformation and dimensionality reduction method, which maps the n-dimensional input data to K-dimensional ($K < n$) linearly uncorrelated variables (mutually independent principal components) with respect to the variability. PCA converts the data by a linear orthogonal transformation using the first few principal components, which usually represent about 80 % of the overall variance. The principal component basis minimizes the mean square error of approximating the data. This linear basis can be obtained by application of an eigendecomposition to the global covariance matrix

estimated from the original data. The characteristic mathematical stages of PCA can be briefly described as follows according to [2], [7]. Firstly suppose that the training data are represented by M n-dimensional feature vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$. One of the integral parts of PCA is the centering of all vectors (subtracting the mean) as:

$$\boldsymbol{\Phi}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}; i \in \langle 1; M \rangle, \tag{1}$$

where:

$$\bar{\boldsymbol{x}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{x}_i, \tag{2}$$

is the mean vector. From the centered vectors $\boldsymbol{\Phi}_i$ the centered data matrix with dimension $n \times M$ is created as:

$$A = [\boldsymbol{\Phi}_1 \boldsymbol{\Phi}_2 \ldots \boldsymbol{\Phi}_M]. \tag{3}$$

To represent the variance of data across different dimensions, the global covariance matrix is computed as:

$$C = \frac{1}{M-1}\sum_{i=1}^{M} \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^T = $$

$$= \frac{1}{M-1}\sum_{i=1}^{M} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T. \tag{4}$$

An eigendecomposition (5) is applied to the covariance matrix in order to obtain its eigenvectors (spectral basis) $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ and their corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, as follows:

$$C\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i; i \in \langle 1; n \rangle. \tag{5}$$

The principal components are represented by the eigenvectors and the most significant ones are determined by K leading eigenvectors resulting from the decomposition. The dimensionality reduction step is performed by keeping only the eigenvectors corresponding to the K largest eigenvalues ($K < n$). These eigenvectors form the transformation matrix U_K with dimension $n \times M$:

$$U_K = [\boldsymbol{u}_1 \boldsymbol{u}_2 \ldots \boldsymbol{u}_K], \tag{6}$$

while $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. Finally, the linear transformation $\Re_n \rightarrow \Re_K$ is computed as:

$$\boldsymbol{y}_i = U_K^T \boldsymbol{\Phi}_i = U_K^T (\boldsymbol{x}_i - \bar{\boldsymbol{x}}), \tag{7}$$

where $\boldsymbol{y}_i$ represents the transformed feature vector. The value of K can be chosen as needed or according to the following comparative criterion:

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{n} \lambda_i} > T, \tag{8}$$

where the threshold $T \in \langle 0{,}9; 0{,}95 \rangle$. T represents the part of the global variance of the original data preserved in the new feature space.

3. Full-Data Trained PCA

In this section, the classical PCA training process is shortly described. At this stage, the whole amount of training data is used. Each parametrized speech signal in the corpus is represented by a separate LMFE matrix. Firstly, the initial data preparation steps are performed. These are described by (1), (2) and (3). The global covariance matrix is computed according to (4) and then decomposed to a set of eigenvector-eigenvalue pairs. According to the K largest eigenvalues, the corresponding eigenvectors were chosen. These formed the transformation matrix U_K (6), which was used to transform the train and test corpus into PCA feature space. Note that the final dimension K of the feature vectors after PCA transformation was chosen to $K = 13$ independently from the criterion formula (8), (because of regular comparison with MFCCs). The new PCA-based corpus was used to train the acoustic model based on full-data trained PCA. This model was created in order to compare the full and partial-data trained PCA models.

3.1. Proposed Method – Eigenvalue Criterion-Based Feature Selection

This section presents three specific algorithms proposed in order to select the most specific feature subset for PCA training. There are two major processing stages. The first one is the "fast" PCA used for feature selection and the second one is the main PCA. The selection approach is based on eigenvalue criterion. The proportion of the first eigenvalue in the eigenspectrum decides whether the analyzed data is significant enough or not. To determine the proportion, following comparative criterion is used:

$$\frac{\lambda_1}{\sum_{i=1}^{N} \lambda_i} > T, \tag{9}$$

where N represents the number of eigenvalues, in this case $N = 26$. The selected data are concatenated into one train matrix, which the input for the main PCA. There are 2 criterion modifications. In case of the first one, if the proportion is greater than T, the analyzed data are stored. The second one stores the data with respect to inversed comparative criterion, that means all analyzed data are stored if the proportion is smaller than T. The data that do not fulfill to the criterion are ignored. The selected data matrix is formed from the most characteristic data for optimal partial PCA training. We propose three feature selection levels based on different algorithms. The first one selects the data on the recording level, the second one analyzes the data on data block level and the third one analyzes the data on feature vector level. The main aspect of proposed algorithms is the training data matrix reduction. Each of the three mentioned algorithms were set to extract data of size 0,05; 0,1; 0,5; 1; 5 and 10 % of the original training set.

1) Recording Level Feature Selection

The recording level selection represents the coarsest method of speech data analysis. The algorithm ignores all those recordings that do not fulfill to the selection condition. However, the ignored recordings could still contain some information rich training data parts. The function of this algorithm illustrates Fig. 1. The parameters for the algorithm are listed in the Tab. 1. In this table Qty (quality) means the amount of the selected subset in percentage.

Fig. 1: Block diagram of the selection algorithm based on recording level analysis.

Tab.1: Parameters for the algorithm based on recording level analysis.

	Normal criterion		Inverse criterion	
Qty [%]	**Threshold/Vectors**		**Threshold/Vectors**	
10,0	0,7625	1909720	0,537	1908732
5,0	0,7955	951559	0,5091	950003
1,0	0,85	193910	0,455	194084
0,5	0,87	94691	0,435	95119
0,1	0,9095	19431	0,4	19237
0,05	0,925	9980	0,39	10042

2) Data block Level Feature Selection

The data block level selection algorithm represents an intermediate level between the three specified approaches. The disadvantage of ignoring of the whole recording due to the recording selection approach is reduced by another dividing of the recording data matrix into smaller blocks. In this work we worked with data blocks with size 26×26. These blocks were the subject of the selection criterion analysis which led to the selected PCA data matrix. The function of this algorithm is illustrated on Fig. 2. The parameters for this algorithm are listed in the Tab. 2.

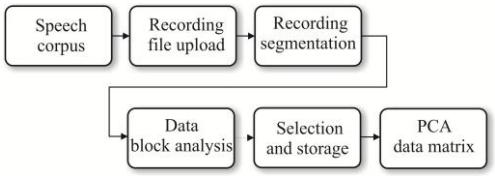

Fig. 2: Block diagram of the selection algorithm based on data block level analysis.

Tab.2: Parameters for the algorithm based on data block level analysis.

	Normal criterion		Inverse criterion	
Qty [%]	**Threshold/Vectors**		**Threshold/Vectors**	
10,0	0,869	1902742	0,5075	1909720
5,0	0,901	954866	0,442	945486
1,0	0,942	192414	0,85	193910
0,5	0,9525	94653	0,2235	94406
0,1	0,968	19135	0,19	18890
0,05	0,9726	9705	0,182	9620

3) Feature Vector Level Selection

The feature vector level selection algorithm stands for the finest method of speech data analysis because each feature vector represents the lowest available data level. The function of this algorithm is similar to Fig. 2 (only the block "Data block analysis" is changed to "Vector analysis").

Data vector level feature selection algorithm operates similarly to the other two mentioned algorithms with the difference at the eigenvalue criterion application. Each LMFE vector is reshaped to matrix in order to compute its covariance matrix, which is treated as the input to the PCA analysis. The parameters for this algorithm are listed in the Tab. 3.

Tab.3: Parameters for the algorithm based on vector level analysis.

	Normal criterion		Inverse criterion	
Qty [%]	**Threshold/Vectors**		**Threshold/Vectors**	
10,0	0,91	1897148	0,629	1906892
5,0	0,9323	951645	0,591	947804
1,0	0,962	190320	0,541	192110
0,5	0,9695	96634	0,529	96226
0,1	0,981	19592	0,513	19138
0,05	0,9844	9543	0,5095	10047

3.2. Experimental Setup

The speech corpus [8] contains approximately 100 hours of spontaneous parliamentary speech recorded from 120 speakers (90 % of men). For acoustic modeling 36917 training utterances were exactly used. For testing purposes, another 884 utterances were used.

The speech was preemphasized and windowed using Hamming window. The window size was set to 25 ms and the step size was 10 ms. Fast Fourier transform was applied to the windowed segments. Mel-filterbank analysis with 26 channels was followed by logarithm application to the linear filter outputs. The 26-dimensional LMFE features were decorrelated by DCT to obtain 13-dimensional MFCC vectors and also used for PCA processing. After PCA, only 13 coefficients were retained. All the MFCC and PCA vectors were finally expanded by delta and acceleration coefficients to 39-dimensional feature vectors.

The acoustic modeling by using HTK Toolkit [9] was performed. The recognition system used context independent monophones modeled using three-state left-to-right HMMs. The number of Gaussian mixtures per state was a power of 2, starting from 1 to 256. The phone segmentation of 45 Slovak phones was obtained from embedded training and automatic phone alignment. During the test, it was used a word lattice created from a bigram language model, which from the test set was built. The vocabulary size was approx. 125*k*. Notice that the accuracies in the evaluation process were computed as the ratio of the number of all word matches to number of reference words.

4. Results and Conclusions

In this paper, we proposed three feature selection algorithms based on eigenvalue-criterion in PCA. Overall 36 experiments were performed. The results are compared to the 39-dimensional reference MFCC model and also to the PCA model (trained from the whole corpus – PCA 100 %). Models were trained for 1−256 Gaussian mixtures. From the Tab. 4 it can be seen that partially trained PCA models achieve comparable or even better results than classical PCA. Accuracies of MFCC model for all mixtures are improved (except 128 mix.) by the proposed method and all accuracies of "PCA 100 %" are improved for all mixtures (italics font in the table). Generally, the best results for 0,05 % part of train corpus for 4 mixtures were achieved (bold marked values). Thus, it is enough a very small amount of speech data to train PCA successfully. We can suppose that the used amount contains probably the most homogeneous data suitable for PCA training. Note that the acoustic models are always trained from the whole corpus so there are enough data to estimate the parameters of Gaussian mixtures. Our proposed method achieves better results at a lower number of Gaussian mixtures (1−8). We suppose better results for higher mixtures in case of a larger amount of speech data. This approach can speed up the PCA training in case of large speech corpora. In the future, we consider the use of different input data kinds for this method and its application to larger speech databases.

Tab.4: Recognition results [%] for the reference MFCC model, PCA model trained from the whole corpus and the partial-data PCA.

Gauss. mix.	Ref. MFCC	PCA (100%)	Partial PCA	Qty [%]	A to MFCC	A to PCA (100%)
1	82,32	82,80	*83,03*	5	+0,71	+0,23
2	83,23	84,10	*85,13*	0,5	+1,90	+1,03
4	85,07	86,01	*87,45*	0,5	+2,38	+1,44
8	87,75	88,88	*89,03*	0,5	+1,28	+0,15
16	89,54	89,84	*90,20*	5	+0,66	+0,36
32	90,84	90,31	*90,92*	0,05	+0,08	+0,61
64	91,41	91,00	*91,54*	0,05	+0,13	+0,54
128	92,34	91,72	*92,26*	0,05	-0,08	+0,54
256	92,51	92,30	*92,62*	0,05	+0,11	+0,32

Acknowledgements

The research presented in this paper was supported by the Ministry of Education of Slovak Republic under research project VEGA 1/0386/12 (50 %) and Research and Development Operational Program funded by the ERDF under the project ITMS-26220120030 (50 %).

References

[1] ABBASIAN, H., B. NASERSHARIF and A. AKBARI. Class-dependent PCA optimization using genetic programming for robust MFCC extraction. In: *Proceedings 3th Conference on Information and Knowledge technology.* Mashad: Ferdowsi University of Mashhad, 2007. Avaible at: http://confbank.um.ac.ir/modules/conf_display/conferences/ikt07/pdf/E1_5.pdf.

[2] JOLLIFFE, Ian. *Principal Component Analysis.* 2nd ed. New York: Springer, 2002. ISBN 03-879-5442-2.

[3] WANG, X. and D. O'SHAUGHNESSY. Improving the efficiency of automatic speech recognition by feature transformation and dimensionality reduction. In: *8th European Conference on Speech Communication and Technology - Proceedings of INTERSPEECH.* Geneva: ISCA, 2003, pp. 1025-1028. ISSN 1018-4074.

[4] HAEB-UMBACH, R. and H. NEY. Linear discriminant analysis for improved large vocabulary continuous speech recognition. In: *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92*. San Francisco: IEEE, 1992, vol. 1, pp. 13-16. ISSN 1520-6149. ISBN 0-7803-0532-9. DOI: 10.1109/ICASSP.1992.225984.

[5] VISZLAY, P., M. PLEVA and J. JUHAR. Dimension reduction with principal component analysis applied to speech supervectors. *Journal of Electrical and Electronics Engineering*. 2011, vol. 4, no. 1, p. 245-250. ISSN 1844-6035.

[6] VISZLAY, P. and J. JUHAR. Feature Selection for Partial Training of Transformation Matrix in PCA. In: *Proceedings of 13th International Conference on Research in Telecommunication Technologies, RTT'11*. Techov: Brno University of Technology, 2011. pp. 233-236. ISBN 978-80-214-4283.

[7] Principal Component Analysis (PCA). In: *University of Montreal* [online]. 2003. Available at: http://www.iro.umontreal.ca/~pift6266/A06/cours/pca.pdf.

[8] DARJAA, S., M. CERNAK, S. BENUS, M. RUSKO, R. SABO and M. TRNKA. Rule-based triphone mapping for acoustic modeling in automatic speech recognition. In: *14th International Conference, TSD 2011*. Berlin: Springer, 2011, lNAI 6836, pp. 268-275. ISSN 0302-9743. ISBN 978-3-642-23537-5. DOI: 10.1007/978-3-642-23538-2_34.

[9] YOUNG, Steve, Gunnar EVERMANN, Mark GALES, Thomas HAIN, Dan KERSHAW, Xunying (Andrew) LIU, Gareth MOORE, Julian ODELL, Dave OLLASON, Dan POVEY, Valtcho VALTCHEV, Phil WOODLAND. The THK Book: for HTK Version 3.4. In: *Cambridge University* [online]. 2009. Available at: http://www.ee.ucla.edu/~weichu/htkbook/.

About Authors

Peter VISZLAY was born in Roznava, Slovak Republic in 1985. He received his M.Sc. in Electronics and Telecommunication from the Technical University of Kosice in 2009. He is a Ph.D. student at the Department of Electronics and Multimedia Communications. His research includes different kinds of speech preprocessing, linear feature transformations and acoustic modeling for Slovak continuous speech recognition systems

Jozef JANECKO was born in Spisska Nova Ves, Slovak Republic in 1989. He received his B.Sc. in Telecommunications from the Technical University of Kosice in 2012. He is an M.Sc. student at the Department of Electronics and Multimedia Communications. He is interested in linear transformations of speech signals.

Jozef JUHAR was born in Poproc, Slovakia in 1956. He

graduated from the Technical University of Kosice in 1980. He received Ph.D. degree in Radioelectronics from Technical University of Kosice in 1991, where he works as an Associate Professor at the Department of Electronics and Multimedia Communications. He is author and co-author of more than 200 scientific papers. His research interests include digital speech and audio processing, speech/speaker identification, speech synthesis, development in spoken dialogue and speech recognition systems in telecommunication networks.